Living With A Star Tools and Methods
Abstracts of Selected Proposals
(NNH21ZDA001N-LWSTM)

Below are the abstracts of proposals selected for funding for the Living With a Star Tools and Methods program. Principal Investigator (PI) name, institution, and proposal title are also included. Thirty-nine proposals were received in response to this opportunity. On January 26, 2022, twelve proposals were selected for funding.

Alex Antunes/Johns Hopkins University
Readying 25 years of full-disk EUV images for machine learning

We will create the full set of ML-ready EUV data from 1995 to present, accessible via the cloud, by bringing in historical restoration of the STEREO/SOHO era in with present datasets into an machine-learning (ML)-ready dataset. The data synthesis will enable 360 degree views of sun and the combined cadences will allow better use of less-sampled EUV instruments with higher-cadence sets. Use of cloud computing will simplify researcher access compared with current high-level datasets. Solving this requires creating homogenous data from source instruments of different resolution, cadence, and slight wavelength differences as well as dealing with cross-calibration. While spatial sampling is a solved problem, differences in cadence in particular will require experimentation to yield cross-comparable results. This work enables research on events, evolution of solar irradiance, segmentation approaches, 360 degree maps of the sun, and other research topics as well as for use with space weather. In terms of data access, at APL we have access to the STEREO/EUVI and SOHO/EIT data, while SDO has created ML-ready data (Galvez et al, 2019). Our team already works with ML and cloud heliophysics data as well as the open standard Heliophysics API (HAPI 2.0) and SunPy.

Graham Barnes/NorthWest Research Associates, Inc
Faster, Better, Deeper: Utilizing Deep Learning to Produce Enhanced Near Real Time Inversions from HMI Data for Space-Weather Modeling.

We propose to use a machine learning (ML) algorithm to simultaneously improve the quality and significantly speed up the production of near real time (NRT) vector magnetic field data from the Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics Observatory. Our general approach will be to use convolutional neural networks (CNN), based on a U-net architecture, combined with regression-by-classification, with Stokes vectors as the input for the network, and various outputs from a Milne-Eddington inversion as the targets that the networks will be trained to produce. This approach has already been trained and validated to provide high fidelity reproductions of the present science quality inversion products of the HMI pipeline at their original resolution (Higgins et al 2021b), as well as to emulate the results of the pipeline inversions of Hinode SOT-SP data (Higgins et al 2021a).

The same type of network will be trained on the NRT Stokes spectra from HMI with SOT-SP inversions as the target. The results will enhance the present NRT HMI pipeline data product by (1) providing the magnetic field estimate for every pixel, including polar regions, (2) reducing systematic bias in the field inclination through an independent estimation of the magnetic fill factor, and (3) estimating the full physical magnetic field vector, i.e., the physically meaningful quantity used as the boundary condition for many space weather models, effectively through an ML-based resolution of the inherent 180 degree ambiguity.

These enhanced data products, produced with substantially reduced processing time, may be useful in space weather forecasting, particularly the real time modeling of eruptions that produce energetic particles, thus giving additional warning for impact on spacecraft throughout the heliosphere. This benefit addresses the third Living With a Star objective, "Human Exploration and Development: LWS provides data and scientific understanding required for advanced warning of energetic particle events that affect the safety of humans." As an additional product, the machine learning-ready data set used to train the network will be provided to the community, ready for similar or hopefully ground-breaking efforts in ML research to improve the inversion and interpretation of Stokes spectropolarimetric spectra.

The CNN to produce near-real-time full-disk vector-field data series will be implemented to run at Stanford University within 18 months of start date, with the output available through the Joint Science Operations Center, the standard repository for HMI and other NASA-mission data. Also within 18 months, the ML-ready SDO-Hinode dataset and corresponding evaluation code will be archived at the University of Michigan Library Deep Blue data repository. This system provides a DOI for the published data, and access via both browser and Globus (which facilitates inter-institutional transfers).

John Dorelli/NASA Goddard Space Flight Center
Vlasov Informed Super Resolution (VISR): A Deep Learning Approach for De-Aliasing Particle Data

Earth's magnetosphere is a complex system with plasma dynamics occurring on a vast range of spatial and temporal scales. Understanding how kinetic plasma dissipation processes shape the large scale structure and dynamics of the magnetosphere remains one of the great challenges of heliophysics. The community's effort to develop the next generation of predictively powerful global space weather models would benefit greatly from observations of plasma dissipation processes on millisecond time scales. NASA's Magnetospheric Multiscale (MMS) mission has paved the way for such observations, but important processes like turbulent dissipation and wave-particle interactions remain marginally resolved at best due to particle instrument aliasing.

We propose to develop a new tool for the heliophysics community -- Vlasov Informed Super Resolution (VISR) -- that uses deep learning to combine time-aliased particle data with the Vlasov equation to recover physically meaningful information below the energy sweep time scale. VISR will make use of Physics Informed Neural Networks (PINN), a recent development in deep learning that has shown promise for a wide range of applications. We will apply VISR to data collected by the Fast Plasma Investigation (FPI) on MMS. After prototyping and validating VISR on test data generated by analytic solutions of the Vlasov equation, we will scale up to production on the ADAPT GPU cluster at NASA-GSFC.

VISR will be made available to MMS team members at the MMS Science Data Center (SDC) no later than 18 months after the start date of the project. After an additional 6 month beta testing period, VISR will be made publicly available under an open source license at both the MMS SDC and an open github repository.

Kiran Jain/ National Solar Observatory
Application of Machine Learning to Improve the Detection of Active Regions in the Sun's Far Hemisphere

Scope and Goal: The measurement of magnetic activity on the solar surface has multiple applications, especially in space weather forecasting. The near-side regions of high magnetic concentration (known as active regions) can be directly observed while for the far-side regions, at present, we mostly rely on indirect methods based on the technique of local helioseismology. In this technique, we map the phase shift (travel time delay) between acoustic waves traveling into a region and its echo. This technique is being used for last two decades and shows/indicates the presence of active regions on the backside of the Sun. However, the detection of the active regions so far is mostly limited to big regions, though there are numerous examples of smaller active regions producing severe space weather events. In the proposed work, we plan to improve the detection of solar active regions on the far-side of the Sun by developing a Machine Learning (ML) tool for application to the helioseismic phase-shift maps. The goal is to lower the threshold of the required strength of the seismic signal for detecting active regions of different sizes. This will be done by reducing the noise and enhancing the seismic signal in farside phase-shift maps.

Methodology and Data: The proposed work is based on the following steps:

(1) Compile a database of GONG helioseismic maps of the Sun's far hemisphere from archives available at the National Solar Observatory (https://nso.edu/data/nisp-data/) concurrent with periods during which NASA STEREO had full EUV coverage of the same. (NSO/GONG has been posting twice-daily synoptic helioseismic maps of the

Sun's far hemisphere at this website since 2006 from 24-hr time series of Dopplergrams. These are archived as open-source datasets (https://farside.nso.edu) and can be easily accessed through the internet.)

(2) Design and train ML-based parameterized algorithms that, applied to the helioseismic phase-shift maps and STEREO-EUV images, propose to recognize and characterize the signatures of active regions.

(3) Formulate an analogous system for validation of the algorithm for application to the helioseismic maps based upon the agreement of its identifications and characterizations derived from the STEREO EUV maps.

(4) Determine the parameter set of the algorithm that optimizes the validation formulated in (3).

(5) Implement the algorithm in the existing pipeline for improved detection of active regions on the farside.

Relevance: The proposed work is directly relevant to one of the NASA's Living With A Star objectives, i.e., ``to quantify the physics, dynamics, and behavior of the sun-Earth system over the 11-year solar cycle'' and contributes significantly to advances in operational space weather forecasting. Our work will improve the use of heliophysics data to immediately benefit users of science data and will enhance the overall experience of the users of science data.

The Archive: We plan to deliver the Machine Learning Trained Active Region Recognition Algorithms (MLTARRA) tool developed in this work and a detailed README file to NASA's Community Coordinate Modeling Center (CCMC) within three months of the completion of this project. The near-real-time forecast and the related data products will be posted at the National Solar Observatory's website (www.nso.edu).

Jeremiah Johnson/University of New Hampshire, Durham
Producing Homogeneous, Machine-Learning Ready Auroral Image Databases Using Unsupervised Learning

Dynamic interactions between solar wind and magnetosphere gives rise to dramatic auroral forms that have been instrumental in the ground-based study of magnetospheric dynamics. Although the general mechanism of aurora types and their large--scale patterns are well-known (Newell et. al. 2009), the morphology of small- to meso-scale auroral forms observed in all-sky imagers and their relation to the magnetospheric dynamics are still in question. A better understanding of the morphology of auroral forms is critical to our understanding of magnetospheric dynamics and the coupling of the magnetosphere to the upper atmosphere. Machine learning offers the possibility of surfacing new knowledge in this area, but most existing auroral image databases are not yet machine learning-ready. A key issue is the lack of ground-truth labels: most widely-used machine learning algorithms are supervised, and without ground-truth labels, cannot be used.

The scientific goal of this project is to deliver a large-scale, homogeneous, machine learning-ready database of auroral images that will enable machine learning-driven investigations into the morphology of auroral forms and the corresponding relationship between these forms and magnetospheric dynamics.

To achieve this goal, we will develop a state-of-the-art unsupervised machine learning algorithm capable of automatically labeling 16 years of white light auroral image data from Time History of Events and Macroscale Interactions during Substorms (THEMIS) ground-based All-Sky Imager (ASI) array. Our approach avoids the necessity of ground-truth labels during training by learning latent representations of the input data that capture inherent structural relationships. These representations can then be used for downstream tasks such as automatic labeling and cluster analysis or investigated in their own right. This work will produce the largest publicly-available, labeled, homogeneous, machine-learning ready auroral image database created to date. It will enable the space science community to conduct statistical studies on the relationship between different categories of auroral images, near-earth solar wind conditions, and geomagnetic disturbances at the earth's surface that were not previously possible. It is relevant to the high-level goal to "Determine the dynamics and coupling of Earth's magnetosphere, ionosphere, and atmosphere and their response to solar and terrestrial input" from the Heliophysics Decadal Survey. The machine learning-ready dataset produced by this research, along with the models and software necessary to reproduce the results, will be delivered to the Space Physics Data Facility on or before the conclusion of the project on October 31, 2023.

Michael Kirk/Atmospheric & Space Technology Research Associates
Discovering Micro Events in AIA using Machine Learning

The catalogue of SDO AIA EUV images is now more than 6 PB and growing every day. Each of these images has 16 million pixels and the database is over 151 million images. Can we search the 2.5 quadrillion pixels to find single pixel eruptions? Does AIA regularly detect coronal 'campfires,' seen by Solar Orbiter? What other small dynamic phenomena are prevalent in the corona?

It is impossible to use conventional methods to analyze such an immense amount of data. This proposal utilizes the cosmic ray spikes database (so-called AIA spikes) and unsupervised machine learning (ML) techniques to search for single pixel eruptions. The cosmic ray spike removal image-cleaning algorithm runs automatically on the level-0 AIA images and removes isolated bright points which are usually cosmic rays but occasionally captures some small bright coronal features. Through using the spikes catalogue, we reduce our data science problem by over three orders of magnitude from a total AIA archive of 1015 pixels to 1012 pixels still a very large data science problem. Through reposing our science question into a data science framework, we further reduce the scope of the parameter space. Previous research has identified that different types of solar eruptions have a distinct evolution when observed in each of the 7 different AIA EUV passbands. For example, a coronal jet regularly has a different peak intensity than a solar flare in the 131 and 171 channels. Thus, passband intensity abstracts events into a 7-dimenstional space where each dimension represents the intensity in an EUV channel.

A recent publication by Young et al. (2021), explores the relationships between the physical properties of AIA spikes in the 171 channel and their coronal environments in 126 cases. They found that 96% of the physical features identified in the AIA spikes have a diameter of less than 2 arcsec, can occur anywhere on the solar disk, and typically evolve over a period of less than 5 minutes, demonstrating that spikes are compact in both time and space. Combining the two spatial and one temporal coordinates with the 7-EUV channel coordinates yields a 10-dimensional space defining a spike.

This work will further develop temporal and spatial filters for linking spikes to their physical origin. Spatial-temporal intersections between spikes in different wavelengths give us insight into how these small features are linked together. Over a two-day sample set, we observe that about 6.5% of the remaining physically interesting spikes occur in all 7 AIA filters. Mapping these spikes back onto the coordinates of the solar surface, there is a concentration in the same location as larger solar features such as coronal holes and active regions. Where individual spikes are coincident in time and location between different AIA wavelengths in this sample set will further refine our physical filters. We eliminated over 99% of the spikes which are genuine cosmic rays using this technique in a two-day demonstration period in 2011.

Methodology and Data

This effort will build a database of every AIA spike in 10-dimensional space (EUV wavelength plus space and time). Abstracting the spikes from its physical origin will allows us to use unsupervised ML to further investigate trends and segment our dataset of over 1010 detections. Utilizing an innovative mix of Bayesian and clustering methods, we will identify clusters of spikes and label commonalities between groups of detections. Once this labeling is complete, we can return to our science question and ask if a group is in fact a 'campfire' observation or some other coronal eruption through extracting a subset of individual spikes, embedding them into a context region, and matching the identified label to the clusters in the database - a process described in Young et al. (2021). This effort delivers a high-value database of AIA spikes that have a coronal origin with labels of likely physical phenomena they represent.

Bennett Maruca/University Of Delaware
Large-Scale Average Trends in Plasma Parameters Across the Heliosphere

Goals and Objectives. The goal of this project is to compile and synthesize historical datasets from in-situ spacecraft into a validated, machine learning (ML)-ready dataset consisting of solar wind parameters spanning the heliosphere. We focus on the fundamental plasma parameters (i.e., density, velocity, temperature, and magnetic field) and those derived therefrom (e.g., plasma beta). Our work has two phases. In Phase 1, we will compile and validate trends in these parameters with distance from the Sun. In Phase 2, we will extend these trends to include not only solar distance but also heliographic latitude. We will publicly release our merged, averaged, and ML-ready data along with our best-fit analytic models for these trends. These analytic models will provide summaries that can be readily utilized for initializing and validating global heliospheric simulations as well as for planning future deep-space missions.

Background. With advances in high-performance computing, global simulations can offer detailed predictions for variations in solar-wind plasma across the heliosphere. Initializing and validating these simulations with data from in-situ spacecraft can be challenging since such measurements are inherently localized. A vast network of in-situ spacecraft spanning the heliosphere would address this issue and revolutionize our understanding of the solar wind, but such a mission is entirely impractical. Nevertheless, many in-situ spacecraft have been sent across the heliosphere. After suitable adjustments for inter-calibration and solar cycle, their data can be taken together to reveal information about the solar wind's large-scale, average expansion -- the background" on top of which transient events (e.g., CMEs and CIRs) develop.

Data and Methodology. We will aggregate publicly available historical datasets from multiple spacecraft, including Parker Solar Probe (PSP), Helios 1 & 2, Mariner 2 & 10, Ulysses, Cassini, Pioneer 10 & 11, New Horizons, and Voyager 1 & 2. Measurements from these spacecraft span three orders of magnitude in solar distance across sixty years. We will pre-average each spacecraft's data to a common cadence. Then, using a well-established technique, we will correct for solar-cycle variations by scaling each average measurement by a time-shifted, contemporaneous, 1-au average value from the OMNI dataset. By binning the scaled, averaged data (first with solar distance and then also with heliographic latitude), we will reveal the large-scale, average trends in the solar wind. For ease of use, we will characterize these trends by using modern machine learning techniques, including non-linear chi-squared minimization and Gaussian processes.

James Mason/University Of Colorado, Boulder
Solar Dynamics Observatory Machine Learning Dataset (SDOML) Improvements

Since the publication of A Machine-learning Data Set Prepared from the NASA Solar Dynamics Observatory Mission" (herein called SDOML) in 2019, there has been a new publication using this dataset every other month. The utility of having the data of all three instruments onboard SDO wrapped up into one uniform package with the data already cleaned is clear. However, as key members of the team that originated this dataset, we have identified improvements that should be made that would further increase the utility, ease of access, and therefore science return. We will focus primarily on 4 new improvements, resulting in the release of SDOMLv2. First, we will repacketize all of the data into the .zarr format. This is a new, well-supported format specifically designed for handling large, multi-dimensional arrays, especially for use in cloud computing. It is the preferred format for platforms such as Pangeo. This will enable faster manipulation of the data for users. Second, we will generate the synthetic SDO/EVE emission lines data product for the entire timespan of the dataset (2010-2021) at full cadence (6 minutes, the SDO/AIA cadence). This method accepts SDO/AIA data as input and has already been demonstrated in case studies. SDO/EVE's 60-360 Å channel ceased functioning in 2014. This new synthetic data restoration will re-enable all of the scientific analyses that this channel previously afforded, for example irradiance coronal dimming studies that this proposal's PI is heavily invested in. Third, we will include the full, cleaned SDO/EVE spectra in the dataset. SDOMLv1 only includes the extracted emission lines product. This will enable more detailed scientific analyses that require the full spectrum, such as the study of Doppler shifts during eruptions. Finally, we will build open source tools and an example gallery to demonstrate the access, manipulation, and some use cases for SDOMLv2, with emphasis on cloud computing.

We will deliver the dataset itself to the Stanford Digital Repository where SDOMLv1 is stored. We will also deliver the dataset to NASA's Solar Data Analysis Center (SDAC), leveraging our team's existing relationship with the team at Goddard Space Flight Center that maintains that resource. Based on SDOMLv1 and the changes described above, we anticipate the dataset will be approximately 10 TB. Delivery preparation will begin in Project Year 2 and be completed by the conclusion of the period of performance (2023-08-31). The example gallery and open source tools will be hosted on GitHub and be publicly accessible through the duration of the proposed effort and afterwards.

SDO is an exemplar of big data and as such is a perfect fit for many existing AI tools. As a flagship, the high profile means that big scientific returns can be expected from modest effort to make the data more easily accessible to those AI tools. Moreover, SDO's status as a flagship means that it has a big community impact when it sets the example for establishing datasets and tools like SDOMLv2, which helps establish such practices as the norm for all missions in heliophysics.

Xing Meng/Jet Propulsion Laboratory
A Retrospective Analysis Toolbox for Ionospheric Total Electron Content Maps

SCIENCE GOALS
Regional and global ionospheric total electron content (TEC) maps have been routinely produced and made publicly accessible by several research institutions around the world. These TEC maps are obtained from Global Navigation Satellite System (GNSS)

measurements using different satellite constellations and selected networks of ground receivers. TEC maps are one of the key ionospheric datasets for understanding height-integrated ionospheric dynamics at various spatial and temporal resolutions. Recently, improvements in TEC data processing over areas of dense ground receiver distributions open possibilities of robust analysis of ionospheric structuring. However, comprehensive analysis of ionospheric structuring over two decades of TEC maps is currently lacking due to large data volume. Corresponding analysis tools are yet to be tailored to apply to the TEC map dataset. Characterizing key features on TEC maps and understanding their dynamic coupling with external drivers can significantly benefit space weather forecasting. The goal of the investigation is to design a software toolbox for analyzing ionospheric TEC maps retrospectively. The toolbox will be able to identify local regions of elevated TEC from TEC maps, extract characteristic features of the TEC enhancement regions, and diagnose solar and interplanetary driving of the TEC intensification. The objectives of the investigation are: 1) Develop existing image-processing programs to extract features of local TEC intensifications on both global and regional TEC maps and 2) Implement transfer entropy to analyze the connection from solar and interplanetary conditions to the extracted features of the local TEC intensifications.

METHODOLOGY
We will address the objectives by building on, improving, and integrating our existing software of feature extraction and transfer entropy. First, we will will extend the existing feature extraction program to output not only the number of TEC intensifications but also characteristics of the TEC intensifications. Second, we will implement a normalized transfer entropy calculator to compute the normalized transfer entropy from the F10.7 and solar wind data to the identified TEC intensifications.This will reveal any non-linear correlation and predictive information transfer from the solar/interplanetary conditions to the TEC intensifications. The feature extraction program and the normalized transfer entropy calculator construct the toolbox. We will test the toolbox for global TEC maps and regional TEC maps of various spatial resolutions during solar maximum and solar minimum years.

DELIVERABLES
Our toolbox will be made available to the community and applicable for analyzing TEC maps produced by any research institution. We will deliver the toolbox, including the feature extraction program and the normalized transfer entropy calculator, to GitHub and to the CCMC as one of the LWS supported tools and methods (https://ccmc.gsfc.nasa.gov/lwsrepository/index.php). The expected delivery date is April and May 2023.

RELEVANCE
The investigation is highly relevant to the LWS Objective 1 Space Science: LWS quantifies the physics, dynamics, and behavior of the sun-Earth system over the 11-year solar cycle.". It also addresses the two of the three overarching science goals of NASA's Heliophysics program: Advance our understanding of the connections that link the Sun, the Earth, planetary space environments, and the outer reaches of our solar system" and

Develop the knowledge and capability to detect and predict extreme conditions in space to protect life and society and to safeguard human and robotic explorers beyond Earth".

Brian Thomas/NASA Goddard Space Flight Center
Using Machine Learning to Detect and Build CME Datasets for Heliophysics

Using Machine Learning to Detect and Build CME Datasets for Heliophysics

We plan to develop a Machine Learning (ML) based solution to identify physical properties of coronal mass ejections (CME). Our approach will provide a long baseline catalog of CME detections for SOHO with derived properties which we can use to better understand the population of CME relative to solar cycle and other physical phenomena. An ML-created catalog solves weaknesses found in other automated catalogs and in manual catalogs. In addition to this catalog of CME events we will also provide the ML models which encapsulate SME expertise for identifying CME. These may be utilized to better detect CME for space weather forecasting and to compare observational data with numerical models and thereby serve to better quantify simulation performance and sidestep human bias in comparison. To perform this work we will utilize computer vision, a field of artificial intelligence that trains computers to interpret and understand images.

We will supply these derived events as a VOEvent formatted file to the Solar Data Analysis Center (SDAC) on or by May 2023. ML models based on YOLO computer vision algorithm and the associated python code for creating and running them will be released as open software on the NASA GitHub (https://github.com/nasa/).

Benoit Tremblay/University Of Colorado, Boulder
Deep Learning for Ensembles of High-Resolution and High-Cadence Transverse Velocity Maps as High-Level Data Products

Knowledge of flows in the solar photosphere is crucial for understanding many aspects of solar physics, from processes in the interior to coronal heating and eruptions. While velocities along the line-of-sight are a common data product derived from spectropolarimetric observations, transverse velocities are not. The aim of this proposal is to provide tools to compute transverse flows from observations and enable science relevant to the Living with a Star (LWS) program, including the study of energy transport from the photosphere to the atmosphere where it can be released in the form of space weather events, and the derivation of realistic boundary conditions for data-driven simulations of the atmosphere necessary for studying these events.

There are three main groups of methods to infer transverse flows from observations. Tracking methods (e.g., Local Correlation Tracking, Balltracking) measure optical flows typically from intensitygrams in the Quiet Sun and magnetograms in active regions.

Physics-based methods use magnetograms and Dopplergrams and solve the magnetic induction equation (e.g., PDFI, DAVE4VM) to infer flows in typically active regions. Finally, the DeepVel & DeepVelU deep learning methods use magnetohydrodynamics (MHD) simulations and a combination of intensitygrams, magnetograms, and Dopplergrams of Quiet Sun or a active region to infer depth-dependent transverse velocities. Through supervised learning, the neural network emulates flows, and by extension the full set of MHD equations, from the training simulation. Recently, we trained DeepVel using simulation data with spatial resolutions and cadences comparable to instruments like SDO/HMI, DST/IBIS, SUNRISE/IMaX, and DKIST/VBI. Our tests using model data have yielded promising results (Tremblay et al., 2021). For example, using DeepVel, we were able to recover flows in the Quiet Sun at subgranular spatial and temporal scales where optical flows decorrelate from physical flows. While these results are encouraging, there remain improvements before DeepVel can confidently be applied to observations. Although simulations have become increasingly realistic, the validity of model-dependent flows inferred through deep learning needs to be addressed thoroughly.

With this proposal, we plan to: (1) Prepare instrument-specific versions of DeepVel trained on an ensemble of state-of-the-art radiative-MHD simulations of granulation, pores, and sunspots to be gathered from the community; (2) Use combinations of inputs (intensitygrams, magnetograms, Dopplergrams) and provide physical interpretations of the results; (3) Improve preprocessing (spatial sampling, PSF convolution, etc.) to project from the model space to the observations space; (4) Assess realism of flows inferred through deep learning by comparing to tracking methods (e.g., FLCT, Balltracking) in weak-field regions and physics-based methods (e.g., PDFI) in strong-field regions; (5) Estimate errors; (6) Experiment with neural network architectures and training, like the development of a physics-informed neural network to generate flows constrained by the induction equation in strong-field regions and by a continuity or transport equation in weak-field regions.

The aim of this proposal is to provide the community with tools to generate realistic transverse velocity maps. We will make the codes available on Github with Jupyter notebook tutorials by 11/01/23. The proposed tools will enhance the scientific output from heliophysics missions (e.g., SDO, Hinode) and target the transport of plasma and magnetic energy, thus addressing LWS objective #1, quantify the physics, dynamics, and behavior of the Sun-Earth system over the 11-year solar cycle", and key goal #4 of the 2013-2022 Heliophysics Decadal Survey Advances in understanding of solar and space physics require the capability to characterize fundamental physical processes that govern how energy and matter are transported."

Jia Yue/Catholic University Of America
Machine learning based automatic detection of upper atmosphere gravity waves from NASA satellite images

Severe weather such as thunderstorms, cold fronts, and hurricanes, excite atmospheric gravity waves (AGWs) that can propagate into the Earth's upper atmosphere. AGWs play key roles in the dynamics and energetics of the mesosphere and thermosphere. They also induce space weather conditions by driving traveling ionosphere disturbances (TIDs) and seeding spread F and plasma bubbles. AGWs imprint their traces in the airglow layers which are several faint emission layers in the mesopause region. AGWs in the upper atmosphere have strong negative correlation with the 11-year solar cycle and the reason is unknown. On the other hand, climate change in the lower atmosphere may change AGW excitations. However, to date, there have been very few satellite observations of global AGWs in the mesopause region. This lack of information about the global AGWs below the E-region ionosphere and lower thermosphere has limited our ability to quantify the impact of AGWs on space weather. Therefore, we propose to undertake a machine learning detection of AGWs in airglow from 10+ years of NASA VIIRS/Day Night Band (DNB) images obtained by two satellites, Suomi NPP and NOAA20 and disseminate the data and algorithm via SPDF. The rich AGW data gained from this work will enable statistical characterization of global AGW morphology in the mesopause region and its solar cycle and long-term variations.
We will build Convolutional Neural Network (CNN) based deep learning models to extract AGW features from DNB images. CNN model is the state-of-the-art technique to classify images and has been widely used in many image detection/classification problems. It typically contains convolutional layer, pooling layer, activation layer such as Rectified Linear Unit (ReLU), fully connected layer, and loss layer in order to capture spatial structure of data in model training. To train CNN models such as 19-layer VGGNet and 50-layer ResNet for AGW detection, we will manually label thousands of images, both with and without AGWs for the training set, and a tenth of the images will be left unused to validate the training model. The training set will be prepared by manually labeling the wave pattern in the wave-containing image. By training the CNN models from the manually labeled images, the models will be able to automatically locate wave patterns and enable us to extract a sub-matrix full of wave patterns from millions of satellite images.

The proposing team combines satellite data processing expertise in both NASA Heliophysics and Earth Sciences, expert in AGW physics, and data science experience in machine learning. The proposed work is listed in the 2021 Heliophysics-LWS Tools announcement: Leverage current technology for the discovery, access, and effective use of NASA's data, as well as enable new technology and analysis techniques for scientific discovery in areas of Heliophysics research covered by LWS objective, quantifies the physics, dynamics, and behavior of the sun-Earth system over the 11-year solar cycle. This work strongly supports the scientific goal emphasized by the 2013-2023 Decadal Survey in Solar and Space Physics, Determine the dynamics and coupling of Earth's magnetosphere, ionosphere, and atmosphere and their response to solar and terrestrial inputs." The machine learning algorithm will also be shared with the NASA WAVE mission.

Deliverables to SPDF by July 2023: ~5000 GW images along with corresponding raw VIIRS data; source codes of image classification and localization models; visualization tools.